Conceptual Mobility and Entrenchment in Introductory Geoscience Courses: New Questions Regarding Physics' and Chemistry's Role in Learning Earth Science Concepts

Steven W. Anderson[1] and Julie C. Libarkin[2]

ABSTRACT

Nationwide pre- and posttesting of introductory courses with the Geoscience Concept Inventory (GCI) shows little gain for many of its questions. Analysis of more than 3,500 tests shows that 22 of the 73 GCI questions had gains of <0.03, and nearly half of these focused on basic physics and chemistry. We also discovered through an assessment of nearly 500 matched pre- and posttests that students were less likely to change answers on basic physics and chemistry questions than they were on those for the geosciences, with many of the low-gain geoscience questions showing switch rates that were similar to that expected for guessing. These results also pertain to the high-scoring pretest students, suggesting that little geoscience conceptual entrenchment occurs for many students enrolled in entry-level courses. Switching rates for physics and chemistry questions were well below the rates associated with geosciences questions, suggesting greater entrenchment. We suggest that students may have difficulty settling on a correct geoscience conception because of the shaky, more entrenched supporting science underpinnings upon which Earth Science ideas are built. These results prompt the following questions: (1) When do our geology majors learn fundamental science concepts if little learning occurs in the introductory courses? (2) What role does the introductory course play in this eventual learning? (3) What strategies can be employed in introductory courses to enhance learning for those students who will only take one college-level geosciences course? We suggest that longitudinal studies of geosciences majors are needed for periods longer than a semester and that more attention be paid to when conceptual change occurs for our majors. © *2016 National Association of Geoscience Teachers.* [DOI: 10.5408/14-017.1]

Key words: concept inventory, geoscience, entrenched ideas, persistent misconceptions, conceptual mobility

INTRODUCTION

Classroom instruction affects the intellectual development of students or student populations in various ways. Student attitudes toward science and scientists, discipline-specific science content, skill development, and conceptual understanding are all student variables that may change as a result of the classroom experience. A reliable measurement of change for any one of these factors would provide instructors with useful information for evaluating the effectiveness of their courses and assessing whether modifications to their instruction are needed. These factors are difficult to evaluate independently, requiring unique assessment instruments that may or may not exist for most disciplines. Thus, most studies of the effectiveness of various educational philosophies and pedagogical approaches rely heavily on anecdotal evidence or qualitative studies that are difficult to generalize, with less widespread quantitative data emerging from use of attitude surveys, concept inventories (i.e., Libarkin, 2008), or similar instruments.

Instructors may view the relative worth of instructional outcomes differently, placing greater emphasis on some outcomes over others depending on their goals for the course. One area that ranks high in terms of relative importance for many college-level faculty members is conceptual understanding. Conceptual understanding implies both a familiarity with content and the ability to apply it to complex questions, and it constitutes some of the more advanced thinking skills important in a college-level education. Over the past 20 years, a number of concept inventories have been developed for determining conceptual change in many science and engineering disciplines, although before 2002, concept inventory development in the Earth Sciences was virtually nonexistent.

In 2002 and 2003, Libarkin and Anderson (2005) administered 29-question and 73-question (respectively) pilot versions of the Geoscience Concept Inventory (GCI v. 1.0; Libarkin et al., 2005; Libarkin and Anderson, 2005) that has subsequently expanded to include more than 100 validated questions (see Libarkin et al., 2011). To ensure external validity and the generalizability of the GCI to entry-level college students nationwide, the 2002 and 2003 administrations of the GCI were completed in 59 courses at 42 institutions across the U.S. Pretesting of 3,595 introductory geoscience students occurred early in the academic year, with posttesting of ~1,750 students collected during the last week of class. At present, the community is invited to use, comment, and add to the GCI through the GCI WebCenter (http://gci.lite.msu.edu; Libarkin et al., 2011).

One interesting trend in the GCI v. 1.0 data that has received little attention is that nearly one-third of the questions show limited, no, or negative pre- to posttest change as a result of college-level instruction across the population of students tested. In other words, the under-

standing of the concepts at the heart of these questions did not significantly improve over the course of a term or semester when measured by the GCI in this population. When measured in individuals, conceptions that do not change despite instruction are referred to as "entrenched" ideas (Vosniadou and Brewer, 1992) or "persistent misconceptions" (Chi et al., 1994). Here, we investigate which Earth Science ideas resist conceptual change in individuals and discuss these in terms of "conceptual mobility" for individual students. In particular, we wish to discover (1) which questions show no or little pre- to posttest change for our entire population of test-takers, (2) whether these questions fall into any particular groups or categories, and (3) whether the students show no or little pre- to posttest change because they were holding firm to incorrect conceptions (entrenchment) or because they were switching from one incorrect answer to another (mobility). We find that even though our test population shows no gain for many questions, individual Earth Science questions record high conceptual mobility and little entrenchment as individual students switch answers over the course of the semester. We also find that basic low-gain physics and chemistry questions show less mobility than the geoscience questions, suggesting that incorrect conceptions are more entrenched in the science disciplines upon which many geoscience concepts are scaffolded. We then discuss the implications this work has for understanding the relationships between teaching and learning in college-level geoscience classrooms.

BACKGROUND

Quantitatively determining the relationships between teaching and learning requires the development of valid and reliable assessment instruments that accurately measure change. Multiple-choice instruments (concept inventories) now exist for assessing conceptual understanding in specific undergraduate science, technology, engineering, and mathematics (STEM) fields, including physics and astronomy (Hestenes et al., 1992; Zeilik et al., 1999; Yeo and Zadnik, 2001; Lindell and Olsen, 2002), chemistry (Krause et al., 2004), geoscience (Libarkin et al., 2005; Libarkin and Anderson, 2007b), and biology and natural selection (Anderson et al., 2002). Allen (2006), Reed-Rhoads (2008), and Libarkin (2008) offer reviews of the state of concept inventory development in STEM disciplines.

Concept Inventory Development—Physics

The physics community pioneered much of the work on developing concept inventories in the physical sciences. Halloun and Hestenes (1985a,b) constructed an open-response exam to measure students' knowledge of mechanics and gave it to more 1,000 students. From these written responses, they selected the most common misconceptions and used them as wrong answers in a multiple-choice test called the Force Concept Inventory (FCI; Hestenes et al., 1992). The FCI was easier to administer and grade compared with open-response items, and it provided access to a large data set for studying the relationships between teaching and learning.

Physics professors and graduate students critiqued early versions of the FCI for clarity. FCI developers also administered the FCI to 11 graduate students, all of whom received perfect scores, and conducted interviews with 22 introductory students to ensure that students clearly understood each question and the possible answers. Finally, exams of 31 A-grade students were analyzed for common misunderstandings that could be attributed to poor question design, and none were found. These measures were implemented to ensure the validity of both the content (content validity) and the clarity (face validity) of each question. Kuder-Richardson tests of results from groups of students examined at different times indicated reliability coefficients of 0.86–0.89, which Hestenes et al. (1992) cited as unusually high values indicative of reliable tests. These pioneering works on mechanics misconceptions inspired a number of more recent studies on student ideas about physics (e.g., Thorton and Sokoloff, 1998; Harrison et al., 1999; Yeo and Zadnik, 2001).

Many researchers have since used FCI data to study relationships between teaching and learning, including work focusing on the quantification of student conceptual change. Hake (1997) proposed that improvement between pretest and posttest is best expressed as normalized gain g (Hovland et al., 1949; Gery, 1972), where

$$g = \text{gain/maximum possible gain}$$

or

$$g = ([\text{posttest}] - [\text{pretest}])/(100 - [\text{pretest}])$$

Thus, if a class averaged 50% on the pretest and 60% on the posttest, then the class-average normalized gain $g = (60\% - 50\%)/(100\% - 50\%) = 0.2$. Gain can be expressed for individual students, performance of a population of students on an entire exam, or populations on specific questions. Although gain is not a statistical measure of effect, it is a useful proxy for considering the influence of instruction on learning.

The GCI

Libarkin et al. (2005) followed many of the test-construction protocols used by the creators of the FCI as they created the GCI v. 1.0. Because of the broad, interdisciplinary nature of Earth Science, they initially restricted their study of student ideas to the following topics: Earth's interior, Earth's crust, and geologic time. A brief questionnaire (Libarkin et al., 2005) was given to 265 students during the 2001–2002 academic year. Student's written responses then inspired the development an interview protocol for studying student misconceptions Libarkin et al. (2005).

Libarkin et al. (2005) selected sites to ensure demographic variability in their study, including a small, private, elite school (Harvard University, HU), a small, state-supported liberal arts college (Black Hills State University, BH), and two large state universities (Indiana University, IU, and the University of Arizona, UA). Interviewers at each study site conducted semistructured interviews; protocol questions guided the initial discussion, and probing questions were used to encourage students to explain responses. Interviews typically consisted of one to four questions and were between 0.5 and 1 hour long (Libarkin et al., 2005). In total, 105 interviews were conducted in 2001 and 2002 (five at HU, 16 at BH, 82 at IU, and two at UA). Libarkin and Anderson (2005) then formulated multiple-choice questions

that used the most common misconceptions from a line of interview questions as wrong answers (distractors). To ensure content validity, the questions, correct answers, and distractors were reviewed by a panel of seven experts and were revised. The final 29 questions that passed the scrutiny of this panel became the GCI 2002 pilot.

The GCI 2002 pilot was administered at the beginning of the fall semester to 2,215 college students in 42 introductory-level courses (including Physical Geology, Historical Geology, Oceanography, and Environmental Science—see Libarkin and Anderson (2005) for a full list of courses and instructional methods) at 32 institutions in 19 states (21 public and six private 4-year institutions, four community colleges, and one tribal college). The pilot was also given to 1,907 students as a semester-end posttest in 30 courses. Individual course enrollments ranged from nine to 210 students, with most courses falling between 35 and 75 students. Faculty instructors for each of tested courses were also encouraged to complete and critique the exam, and 21 faculty members participated (Libarkin et al., 2005; Libarkin and Anderson, 2005). Instructors also provided a self-report of their estimates of the time spent on each of a variety of instructional strategies. Teaching approaches varied greatly, such that the reported percentage of class time devoted to lecture ranged from 0% to 100%, demonstration ranged from 0% to 30%, small-group work ranged from 0% to 50%, lab exercises ranged from 0% to 60%, and use of technology ranged from 0% to 100%, although faculty self-reporting of teaching approaches is probably less accurate than direct classroom observation (e.g., Johnson and Roellke, 1999).

The GCI was expanded in 2003 to a total of 76 questions, including the 29 questions from the GCI 2002 pilot. The scope of the exam was broadened to include questions on basic physics and chemistry, as well as an expansion of Earth Science topics. Two tests, one with 29 questions and another with 30 questions, were piloted in 2003 as pre- and posttests. Each of these tests contained six common items drawn from the 2002 pilot and 47 new questions divided between the two exams.

The database for the 76 questions from the 2003 GCI pilot contained the responses from 3,595 students who took either the 2002 or the 2003 pilot exams. Individual questions were answered by as many as 3,595 students and as few as 306 students. The large sample size enabled statistical validation of the 2003 GCI pilot and allowed for a study of the relationships between conceptual change and student demographic data, institution characteristics (type, class size, and location), and teaching style. Calibration of item difficulty estimates was performed using Quest software (Adams and Khoo, 1996), the one-parameter logistic model for Rasch analysis, and the Mantel-Haenszel approximation of differential item functioning (DIF) for all 76 questions. Of the 76 questions on the 2003 pilot, two were phenomono-graphical and are no longer used for GCI testing. A third item was removed because of faculty concerns about the accuracy of the item stem, bringing the total remaining questions to 73.

To ensure that pre- to posttest scores reflect the degree to which college students learn, and not the result of a flawed assessment instrument, the creation of the GCI involved validity and reliability measures that went beyond those employed for concept inventory development in other scientific disciplines (Libarkin and Anderson, 2007a,b). The

GCI was created through a multistep methodology, combining scale development theory, grounded theory, and item response theory, incorporating a mixed methods approach and using advanced psychometric techniques not commonly employed in developing content-specific assessment instruments (Libarkin et al. 2005; Libarkin and Anderson, 2007a,b).

For example, one of the most important factors in creating a valid and reliable multiple-choice exam is ensuring that all potential wrong answers, or distractors, are attractive alternative answers to some segment of the test-taking population. If the instrument lacks these attractive alternative distractors, then students may chose the correct answer simply because they did not find the distractors reasonable. To ensure that the wrong answers in the GCI were all attractive to an introductory college student population, incorrect answers that appeared multiple times in the 105 interviews conducted during the early stages of GCI creation were crafted into distractors. In addition, the correct answer must be written in language that introductory students can understand yet pass scrutiny of a panel of experts. Student interviews were again mined for language when constructing the correct answer, and then a panel of geoscience and education professionals commented on whether it was scientifically accurate (Fig. 1). Another important factor considered during GCI construction was determining whether the test was too hard, or too easy, for this population of students. If the concept inventory is too difficult, then students may answer most or all questions incorrectly even though they may have some understanding of the material. If the concept inventory is too easy, then the pretest scores may be too high to show additional improvement on the posttest. Rasch analysis determined the difficulty of individual items and the test to ensure that the test was capable of capturing learning when administered to this population of students. Therefore, this instrument was validated specifically for college students, not for other learning groups, and the results of studies using GCI data cannot be reliably extended to noncollege groups. Specifically, Table I summarizes all validity and reliability measures that went in the construction of the GCI (from Libarkin and Anderson, 2007a), ensuring a valid and reliable tool for measuring pre- to posttest gains in this population of students.

Entrenched Ideas and Persistent Misconceptions

Earth Science ideas that are resistant to conceptual change, also referred to as entrenched ideas or persistent misconceptions, are addressed in a few studies of elementary school-age children. Vosniadou and Brewer (1992) identified two dominant entrenched ideas about Earth: that ground is flat and that all things, including Earth, fall downward. Gordon (1992a,b) studied conceptual change in a single 6th grader and found that conceptual change did not occur quickly, similar to the results discussed by Bruer (1993) and Gardner (1991). Maria (1997) tracked conceptual change in a single boy between kindergarten and 2nd grade and found that it took more than a year for him to restructure an entrenched conception of gravity despite instruction, consistent with other studies that have found that conceptual change proceeds at a slow pace (Gardner, 1991; Gordon, 1992a,b; Bruer, 1993).

a. The following maps show the position of the Earth's continents and oceans. The ♂'s on each map mark the locations where earthquakes occur most frequently. Which map do you think best represents where earthquakes occur most frequently on Earth?

Circle one: **A   B   C   D   E**
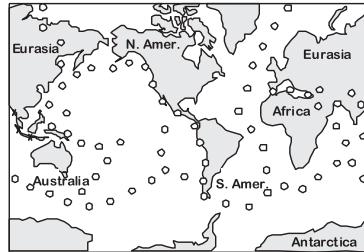
A. In continental and oceanic crust, and along continental margins
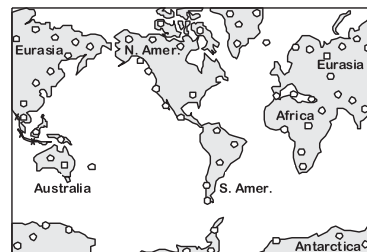
B. Mostly in continental crust
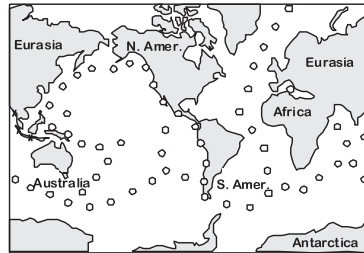
C. Mostly in oceanic crust
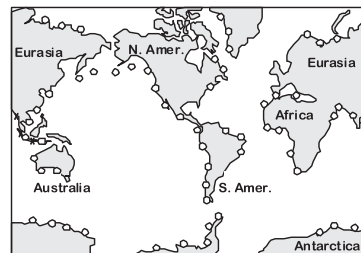
D. Mostly along continental margins
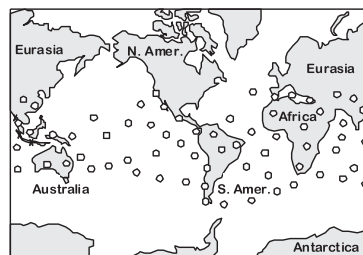
E. Mostly in warm climates

FIGURE 1: Sample GCI question. All of the maps used as distractors originated from student drawings constructed during interviews of 105 students (Libarkin et al., 2005). The correct answer (A) passed scrutiny of a panel of geoscience and education experts.

METHODS

To identify the degree to which ideas are entrenched in the geosciences, we first determined which test items showed no improvement for our pilot-tested population despite instruction. Although there are several potential methods for parsing our data to a "no improvement" subset, we used a simple, straightforward approach and compared the normalized gain for all 73 test items in the GCI v. 1.0. One advantage of using normalized gain is that it reveals both the easy and the difficult test items that show little or no improvement over the course of instruction. We

considered the concepts tested by these no-, low-, or negative-gain questions to be potentially entrenched. A lack of gain may reflect ideas resistant to change (entrenched). However, students may not be showing any improvement because they are switching between different incorrect answers from the pre- to the posttest, thus indicating much mobility in their thinking and little entrenchment.

To identify which of these low-, no-, or negative-gain questions truly represent entrenchment, we tallied how many students chose a particular distractor on the pretest and compared that number with choices on the posttest

TABLE I: Validity and reliability measures used in developing the GCI (from Libarkin and Anderson, 2007a).

Validity/Reliability*	Exemplar Question	Example of Method Used for GCI Development
Construct validity	Is there strong support for content of items?	1) Multimethod: GCI stems and items based upon a large interview data set (n = 75) and questionnaires (n = 1,000); items developed naturally from data (grounded); think-aloud interviews with students 2) Multitrait: Each concept covered by multiple questions
Content (face) validity	Do items actually measure conceptions related to "geoscience"?	1) Review of each question by 6–10 geologists or science educators 2) Review of revised items by 10–21 faculty members for content and correctness of responses
Criterion validity	Correlation between GCI and other measures?	1) Trends in quantitative GCI data correlate strongly with conceptions revealed in qualitative data 2) Preliminary GCI 15-item subtest results show correlation between subtests
External validity	Are results generalizable to other populations?	1) Piloting with wide range of students from 49 institutions 2) Calculation of bias relative to gender and/or ethnicity of subjects via DIF; caution with 4 items suggested by Mantel-Haenszel DIF approximation
Internal validity	Random sample? Do researcher expectations or actions bias results?	1) Items reviewed by experts in both geology and education 2) GCI administered by participating faculty; no administration bias on the part of GCI developers 3) Rasch scales similar for pre- and posttests, suggesting that student attrition and changes made to items during revision do not affect the stability of questions on the Rasch scale
Reliability (repeatability)	One example: Are test results repeatable?	1) Administration to multiple populations yielded similar results 2) Classical reliability and Rasch scale stability 3) Internal consistency of items (KR-20) = 0.69 4) Item separation reliability of Rasch scale = 0.99

(Table II). Movement by the group toward, or away from, one or more distractors would indicate mobility in their conception, not entrenchment. For those questions that exhibited no significant movement between incorrect responses, we looked at the answers of each student to be certain there was no equal movement between incorrect distractors that may indicate guessing. We also determined which general concepts were tested with each question. This methodology allows us to pinpoint the questions and associated concepts in which individual students chose a particular response and stayed with it despite instruction.

To determine whether students are switching their answers from pre- to posttest, we identified those students who took the same version both pre- and posttest. We started with the 2002/2003 database of the 73 GCI v. 1.0 questions and used demographic data to identify all individuals who took the same version to determine how their responses changed as a result of instruction. From the original database of 3,595 students, we found 392 and 102 individuals who completed the same pre- and posttest version of the 2002 and 2003 GCI pilot, respectively. This left us with sample sizes ranging from 48 to 392 individuals who answered the same test items on both pre- and posttests. Table II shows the question number and gain for all 73 GCI v. 1.0 questions. Table III shows each of the 22 low-, no-, or negative-gain questions; a brief description of the concept tested; the number of students who answered the question

on matching pre- and posttest versions; the normalized gain; and a comparison of student responses pre- to posttest.

RESULTS

Several results from this study are evident in Table II:

1. The normalized gain for each of the 73 GCI v. 1.0 items ranged from −0.31 to +0.48 for the entire pilot-tested population, within the range of low to medium normalized gain as described in Hake (1997).
2. Of the 73 questions, 22 had normalized gains of 0.03 or less. Although this boundary is arbitrary, the lack of significant positive change on these questions suggests that the concepts tested are potentially entrenched within this test population.

For these 22 low-gain questions, we also tabulated the number of students who chose each incorrect distractor on the pre- and posttest and determined whether any systematic movement toward a particular wrong answer might be indicative of concept mobility rather than entrenchment. Table III shows each of the 22 questions and the number of pre- and posttest responses, as well as the percentage of students who switched their answers pre- to posttest. We include a column showing the percentage of students who answered the question correctly on the pretest but changed

TABLE II: Gain and statistics for all GCI questions.

Question Number (Test Version)	Gain	Correct on Posttest	Correct on Pretest	Total Matched Pre- and Posttest Questions
2003 Q21 (a)	−0.31	16	25	54
2003 Q14 (b)	−0.15	18	22	48
2003 Q17 (b)	−0.14	6	11	48
2003 Q3 (a)	−0.09	17	20	54
2003 Q12 (a)	−0.07	9	12	54
2003 Q29 (b)	−0.06	11	13	48
2003 Q7 (a)	−0.04	2	4	54
2003 Q15 (a)	−0.04	27	28	54
2003 Q24 (a+b)	−0.03	64	65	102
2003 Q8 (b)	−0.02	6	7	48
2003 Q27 (a)	−0.02	2	3	54
2003 Q8 (a)	0.00	4	4	54
2003 Q13 (a)	0.00	33	33	54
2003 Q30 (a)	0.00	7	7	54
2003 Q4 (b)	0.00	7	7	48
2003 Q28 (b)	0.00	20	20	48
2003 Q27 (b)	0.03	13	12	48
2003 Q3 (b)	0.03	17	16	48
2003 Q12 (b)	0.04	23	22	48
2003 Q7 (b)	0.04	3	1	48
2003 Q13 (b)	0.05	28	27	48
2003 Q4 (a)	0.05	18	16	54
2003 Q17 (a)	0.06	5	2	54
2003 Q18 (b)	0.06	17	15	48
2003 Q29 (a)	0.06	24	22	54
2003 Q6 (a)	0.09	13	9	54
2003 Q16 (a)	0.10	9	4	54
2003 Q2 (a+b)	0.12	17	5	102
2003 Q21 (b)	0.13	22	18	48
2003 Q20 (a)	0.14	29	25	54
2003 Q5 (b)	0.14	36	34	48
2003 Q15 (b)	0.15	26	22	48
2003 Q6 (b)	0.16	21	16	48
2003 Q9 (b)	0.16	32	29	48
2003 Q16 (b)	0.16	22	17	48
2003 Q5 (a)	0.19	32	27	54
2003 Q10 (a)	0.19	32	27	54
2003 Q9 (a)	0.20	42	39	54
2003 Q22 (a+b)	0.20	47	33	102
2003 Q10 (b)	0.21	17	9	48
2003 Q25 (a+b)	0.23	33	12	102
2003 Q19 (b)	0.24	23	15	48
2003 Q28 (a)	0.26	29	20	54
2003 Q11 (a)	0.27	30	21	54
2003 Q11 (b)	0.28	35	30	48

TABLE II: continued.

Question Number (Test Version)	Gain	Correct on Posttest	Correct on Pretest	Total Matched Pre- and Posttest Questions
2003 Q19 (a)	0.28	26	15	54
2003 Q26 (a+b)	0.28	41	17	102
2003 Q23 (a+b)	0.29	72	60	102
2003 Q20 (b)	0.30	29	21	48
2003 Q14 (a)	0.31	36	28	54
2003 Q18 (a)	0.48	38	23	54
2002 Q7 (a+b)	−0.20	53	109	392
2002 Q11 (b)	−0.03	72	75	193
2002 Q20 (a)	−0.01	61	62	199
2002 Q17 (a+b)	0.03	27	15	392
2002 Q5 (a+b)	0.03	142	133	392
2002 Q15 (a+b)	0.06	98	78	392
2002 Q4 (a+b)	0.09	222	205	392
2002 Q6 (a+b)	0.09	242	227	392
2002 Q11 (a)	0.10	40	23	199
2002 Q10 (a)	0.10	78	65	199
2002 Q20 (b)	0.10	57	42	193
2002 Q14 (a)	0.11	115	105	199
2002 Q1 (a+b)	0.12	165	134	392
2002 Q9 (b)	0.15	100	84	193
2002 Q19 (a)	0.16	80	58	199
2002 Q8 (a)	0.18	152	142	199
2002 Q19 (b)	0.19	94	71	193
2002 Q18 (a+b)	0.20	263	231	392
2002 Q16 (a+b)	0.20	228	187	392
2002 Q2 (a+b)	0.20	251	215	392
2002 Q14 (b)	0.22	114	92	193
2002 Q10 (b)	0.24	88	54	193
2002 Q9 (a)	0.27	91	52	199
2002 Q3 (a+b)	0.29	228	160	392
2002 Q13 (a)	0.29	175	165	199
2002 Q12 (a)	0.32	100	54	199
2002 Q12 (b)	0.44	174	159	193
2002 Q13 (b)	0.46	173	156	193
2002 Q8 (b)	0.47	158	127	193

to an incorrect item on the posttest, as well as the difficulty of each question from Rasch analysis (Libarkin and Anderson, 2007b). If all students were simply guessing, we would see 75% change their answers from pre- to posttest for four-item multiple-choice questions and 80% change their answers for five-item questions. We would also expect to see 75%–80% of the students change from the correct answer on the pretest (depending on whether the question is four or five items) to an incorrect posttest response.

When we looked only at the low-gain subset of 22 questions in Table III, we found the following:

1. Of the 22 low-gain questions, nine (40.1%) tested basic physics and chemistry principles (less than 25% of all GCI questions, or 17 of 73 questions, were chemistry and physics); two were related to general geology principles; three covered Earth size, shape, or origin; two focused on geologic time; three dealt with erosion; two covered volcano, tectonics, and earthquakes topics; and one related to the atmosphere.

2. The highest percentage of answer switching was 87.2% (for a pilot eight-item question on techniques

TABLE III: Gain and statistics for 22 low-gain questions.

Physics

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Paraphrased Questions Why did aluminum ball in #8 above behave	Changed to on Post
10	0	54	4	4	37	17	31.5	1.09	2	2	2	50	A	33	A	33	magnetic field significantly affects path	7
													B	12	B	10	magnetic field slight affects path	5
													C	4	C	4	magnetic field does not affect path	2
													D	0	D	1	magnetic field makes a planet heavier	1
													E	6	E	6	magnetic field affects path and makes a planet heavier	3

Pre²
```
   A  B  C  D  E  NR
Post A 27  3  2  0  2  0
     B  4  5  0  0  1  0
     C  0  2  2  0  0  0
     D  0  1  0  0  0  0
     E  2  1  0  0  3  0
    NR  0  0  0  0  0  0
```

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Fate of a wooden satellite in orbit around the Earth	Changed to on Post
31	0.03	48	16	17	23	25	52.1	0.17	8	7	9	43.8	A	6	A	10	Continue to orbit because of the magnetic field	7
													B	16	B	17	Continue to orbit because of gravity	8
													C	16	C	21	Would float into space because magnetism does not work on wood	10
													D	7	D	0	Fall towards the Earth because the magnetic field would have no effect	0
													E	3	E	0	Fall towards the Earth because gravity would not have any effect	0

Pre
```
   A  B  C  D  E  NR
Post A  3  2  2  2  1  0
     B  3  9  3  2  0  0
     C  0  5 11  3  2  0
     D  0  0  0  0  0  0
     E  0  0  0  0  0  0
    NR  0  0  0  0  0  0
```

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Drop a rock into a tunnel cutting through the Earth at the Equator	Changed to on Post
52	-0.15	48	18	18	30	29	62.5	-0.66	9	17	12	77.3	A	5	A	3	Rock would fall south	3
			4 students										B	3	B	8	Rock would fall completely through the Earth and stop at the other side	7
			chose										C	6	C	7	Rock would fall completely through the Earth and keep going	5
			mult answers										D	15	D	13	Rock would fall into the hole, but drift south and stop	10
			including										E	22	E	18	Rock would fall to the center of the Earth	9
			correct															
			distractor															

Pre
```
   A  B  C  D  E  NR
Post A  0  1  0  2  0  0
     B  1  1  1  3  2  1
     C  1  1  3  1  2  0
     D  2  0  1  4  7  0
     E  1  0  1  6 12  1
    NR  0  0  0  0  6  0
```

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Aluminum ball passing between 2 planets - one magnetic	Changed to on Post
8	-0.04	54	4	2	25	29	53.7	1.35	1	3	1	75	A	6	A	6	curve and hit magnetic planet	5
													B	18	B	20	curve around magnetic planet	9
													C	5	C	8	curve away from magnetic planet	6
													D	22	D	19	orbit magnetic planet	5
													E	4	E	2	pass between two planets	1

Pre
```
   A  B  C  D  E  NR
Post A  2  2  0  2  1  0
     B  2 11  1  5  1  0
     C  1  2  2  2  1  0
     D  1  2  2 14  0  0
     E  0  1  0  0  1  0
    NR  0  0  0  0  0  0
```

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		What happens if you drop a steel ball in North America?	Changed to on Post
54	-0.04	54	28	27	27	27	50	-0.89	8	10	19	35.7	A	8	A	11	Roll towards equator	8
			(2 chose										B	5	B	11	Roll north	8
			mult answers			Average Change in Group	41.6						C	1	C	2	Roll towards nearest ocean	2
			including						Average change in Group		47.0		D	28	D	27	Roll downhill	8
			correct										E	12	E	6	Roll towards nearest magnet	4
			distractor															

Pre
```
   A  B  C  D  E  NR
Post A  3  0  0  5  3  0
     B  4  3  0  2  2  0
     C  0  1  0  1  0  0
     D  0  1  1 19  6  0
     E  3  0  0  1  2  0
    NR  0  0  0  1  5  0
```

Chemistry

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Can you determine if there is iron in a black rock?	Changed to on Post
29	-0.07	54	12	9	19	35	64.8	0.33	8	11	19	91.7	A	0	A	3	Yes, black rocks contain iron	3
													B	9	B	16	Yes, there will be silver specks if there is iron	13
													C	28	C	22	Yes, you could see the iron with a microscope	9
													D	12	D	9	No, if present you couldn't see iron even with a microscope	8
													E	5	E	4	No, the rock is invisible if it doesn't reflect light.	3

Pre
```
   A  B  C  D  E  NR
Post A  0  0  1  2  0  0
     B  0  3  6  5  2  0
     C  0  6 13  3  0  0
     D  0  0  7  1  1  0
     E  0  0  1  1  2  0
    NR  0  0  0  0  1  0
```

for calculating the age of Earth that allowed for multiple correct answers).

3. Of the four- and five-item questions that allowed a single correct answer, the highest percentage of switching found was 81.5% for a five-item geoscience question on the relationship between mountain morphology and time.

4. The lowest switching rate for these 22 low-gain questions was 31.5% on a five-item question dealing with the difference between gravity and magnetism.

We were also interested in the switching rates for those students who correctly answered the pretest question—did they stay with their correct answer on the posttest, or did they switch as well? We found the following:

1. The highest rate of switching for correct pretesters was 85.7% for a five-item question on the definition of a tectonic plate (only one student kept the correct answer pre- to posttest).

2. The lowest rates of switching for correct pretesters were on three questions dealing with the size and shape of Earth, with 24.2% (four-item question,

25.0% (five-item question), and 29.2% (three-item question) switching.

One trend is that half of the 22 questions had switching rates that were within 20 percentage points of what would be expected if the students were guessing, suggesting that conceptual entrenchment is not occurring with the concepts at the heart of these questions. However, some questions, such as the three questions dealing with the size and shape of Earth, have switching rates that cannot be explained by guessing alone.

Of interest is that six of the eight physics and chemistry questions showed the lowest overall switching rates. Only two groups of questions showed less than 50% of students switching answers from pre- to posttest; the five physics questions showed the least overall switching from pre- to posttest (41.6% of students switched their answers), as did the three questions on the size and shape of Earth (44.1% switching). One of the three questions on the size and shape of Earth had only three items, and less switching through guessing is expected simply because of fewer choices for that question. Therefore, only the physics questions appear anomalous in terms of having a lower percentage of

TABLE III: continued.

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Place a magnet next to a dull black that isn't magnetic	Changed to on Post		Pre
11	-0.02	48	7	6	22	26	54.2	1.03	4	5	2	71.4	A	7	A	6	Iron could be present because some black rocks have iron	4		A B C D E NR
													B	1	B	3	Iron is present because all black rocks contain iron	3	Post A	2 0 3 1 0 0
													C	17	C	18	No metals present because metals are magnetic	7	B	0 0 1 2 0 0
													D	17	D	9	Metals could be present, but not iron. Iron is red	3	C	3 0 11 3 1 0
													E	5	E	12	No metals. Metals are shiny	9	D	0 0 2 6 1 0
																			E	2 1 0 5 3 1
																			NR	0 0 0 0 1 0

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Appearance of Aluminum in Rock	Changed to on Post		Pre
12	-0.14	48	11	6	20	28	58.3	1.07	3	8	3	72.7	A	15	A	22	Visible Al pieces with other pieces in rock	13		A B C D E NR
													B	4	B	3	Visible balls or sheets unattached to rock that can be picked up	2	Post A	9 3 4 5 0 1
													C	15	C	17	Could see Al with a microscope	11	B	0 1 2 0 0 0
													D	11	D	6	Probably couldn't see Al, even with a microscope	3	C	4 1 7 3 1 2
													E	1	E	1	Not a choice - but one student choose it anyway	1	D	1 0 2 3 0 0
																			E	1 0 0 0 0 0
																			NR	0 0 0 0 0 0

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		True statements about radioactivity radioactivity	Changed to on Post1	
1	-0.02	199	63	60	62	137	68.8	3.73	24	27	36	42.9	A	48	A	53	Occurs only if carbon is present	26	Footnote 3
													B	25	B	21	Cannot occur at the Earth's surface	15	
													C	49	C	43	Is created by people	27	
			Average change in Group			61.5	Average change in Group		69.7				D	113	D	130	half life is a measure of how fast radioactivity decreases	40	
													E	36	E	48	half life decays and eventually disappears	32	

Earth Size Shape, Origin

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Why will the Earth always be, or cease to become, a planet or cease to become, a planet	Changed to on Post		Pre
phenom Q	0	54	8	8	29	25	46.3	-0.67	4	3	6	37.5	A	13	A	9	The Earth has been a planet a long time	7		A B C D E NR
													B	10	B	13	Earth will find a way to continue	13	Post A	6 4 2 0 1 0
													C	16	C	7	Life will ensure the planet's survival	6	B	4 4 5 0 3 1
													D	8	D	8	Earth will become part of something else	4	C	3 2 3 1 0 0
													E	16	E	19	Nothing lasts forever	19	D	1 1 2 6 0 0
																			E	2 2 3 2 12 0
																			NR	0 0 1 0 0 0

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		How big was the Earth when dinosaurs appeared	Changed to on Post		Pre
63	0	54	33	33	33	21	38.9	-1.16	8	8	25	24.2	A	5	A	6	Smaller than today	4		A B C D NR
													B	6	B	6	Larger than today	4	Post A	2 1 2 1 0
													C	33	C	33	Same as today	8	B	1 2 2 1 0
													D	10	D	9	No way of knowing	5	C	2 2 25 4 0
																			D	0 1 4 4 0
																			NR	0 0 0 0 0

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Size of Earth at Formation	Changed to on Post		Pre
44	-0.03	102	65	64	54	48	47.1	-0.42	19	22	45	33.8	A	12	A	11	Smaller than today	9		A B C D NR
		1 student chose mult answers including correct distractor	1 student chose mult answers including correct distractor		Average Change in Group								B	21	B	25	Larger than today	18	Post A	2 3 6 0 0
							44.1	Average change in Group		31.9			C	65	C	64	Same as today	19	B	3 7 13 0 2
													D	1	D	5	Not a choice, but some students chose it anyway	5	C	6 10 45 1 2
																			D	1 1 3 0 0
																			NR	0 1 0 0 0

switching than did other question subsets, possibly indicating more conceptual entrenchment within the test population.

Finally, we were interested in whether students chose particular correct or incorrect answers on the posttest items. In other words, are students moving toward any particular concept (either correct or incorrect) as a result of instruction that might indicate some type of conceptual change? We found that only two questions show a strong move toward one distractor (more than 2:1 over the next most-chosen item). The first question of this type asked students to choose a diagram that best depicted where volcanoes are located on Earth. This question had a correct answer (a diagram showing volcanoes along convergent and divergent margins worldwide) and five additional distractors (one showing volcanoes along most coastlines, including the Atlantic; one with volcanoes only along the Atlantic; one with volcanoes only in warm climates; one with volcanoes mostly on continents; and one with volcanoes mostly on islands). For this particular question on volcano distribution, we found the following:

1. Approximately 70% of all students (n = 267 of 392) switched their answers pre- to posttest, including nearly 83% of the correct pretesters.

2. Of the 267 students who switched to a different answer on the posttest, nearly 45% (119) chose the answer showing volcanoes along all coastlines, including the Atlantic. The next most-chosen distractor was the warm-climate option.

The choice of a warm climate option aligns with interview evidence showing students incorrectly believe that there is relationship between warm atmospheric temperatures and volcanic eruptions (e.g., Libarkin, 2006), although we are puzzled by the prevalence of the coastline misconception. Only 19 of the 109 students (17.4%) who answered the question correctly on the pretest kept the correct answer on the posttest, and only 33 of the 291 students (11.3%) who changed their answers chose the correct response option on the posttest. Similarly, we note a strong move on an item dealing with the location of cloud formation from the correct answer (over oceans) to an incorrect one (equator; 10 of the 20 students who answered correctly in the pretest changed their answer to the equator option on the posttest, and 24 of the 54 total students chose this option on the posttest). We do not have interview data, or information from previous studies, that provide us with a basis for interpreting why students are moving from a correct answer to that particular distractor.

TABLE III: continued.

Erosion

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		If sand blows across the ocean, what will the ocean look like	Changed to on Post
47	-0.31	54	25	16	25	29	53.7	-0.51	3	15	13	60	A	14	A	22	Flat island of sand	12
													B	6	B	3	Mountain of sand	1
													C	5	C	7	Flat island of rock	7
													D	4	D	6	Mountain of rock	6
													E	*25*	*E*	*16*	Unchanged	3

Pre / Post matrix (Question 47):

	Pre A	B	C	D	E	NR
Post A	10	1	3	2	6	0
B	0	2	0	1	0	0
C	1	2	0	1	3	0
D	0	1	2	0	3	0
E	3	0	0	0	13	0
NR	0	0	0	0	3	0

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Which of the following affects erosion rates affects erosion rates	Changed to on Post1
2	-0.02	54	2	1	20	34	63.0	2.95	1	2	0	100	A	41	A	40	Rock type	5
													B	33	B	26	Earthquakes	9
													C	40	C	39	Time	8
													D	*48*	*D*	*45*	Climate	4

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Caused by Wind	Changed to on Post1
45	0	48	20	20	15	33	68.8	-0.35	11	11	9	55	A	6	A	7	Plate movement	6
													B	*32*	*B*	*32*	Waves	21
													C	4	C	6	Earthquakes	5
													D	7	D	10	Mountain building	6
													E	*32*	*E*	*35*	Erosion	23

Average change in Group 61.8   Average change in Group 71.7

Volcano/Tect/EQ

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Definition of a Tectonic Plate	Changed to on Post
6	0	48	7	7	17	31	64.6	1.83	6	6	1	85.7	A	4	A	4	All solid rock beneath the continents/above moving rock	4
													B	28	B	19	All solid rock beneath continents and oceans/above moving rock	8
													C	4	C	10	Solid rock beneath the loose dirt/above moving rock	8
													D	*7*	*D*	*7*	All solid rock and dirt above moving rock	6
													E	5	E	8	Rigid material of the outer core	6

Pre / Post matrix (Question 6):

	Pre A	B	C	D	E	NR
Post A	0	3	0	0	1	0
B	3	11	0	4	1	0
C	1	7	2	0	0	0
D	0	6	0	1	0	0
E	0	1	2	2	3	0
NR	0	0	0	0	0	0

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Location of Volcanoes	Changed to on Post
13	0.02	392	61	69	101	291	74.2	1.13	52	46	15	75.4	A	106	A	185	Pacific and Atlantic margin	119
													B	*109*	*B*	*52*	Pacific margin	33
													C	14	C	7	Atlantic margin	7
													D	61	D	69	warm climates	52
													E	20	E	14	mostly on continents	13
													F	76	F	60	mostly on islands	35

Pre / Post matrix (Question 13):

	Pre A	B	C	D	E	F	NR
Post A	66	58	2	27	6	22	3
B	12	19	3	4	2	11	1
C	4	2	0	0	0	1	0
D	12	16	5	15	6	13	0
E	4	2	1	2	1	3	1
F	7	10	3	12	3	25	0
NR	1	1	0	1	8	1	0

Average change in Group 69.4   Average change in Group 80.6

General Geology

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Origin of Geothermal Heat	Changed to on Post1
27	0	48	11	11	16	32	66.7	0.43	4	4	7	36.4	A	5	A	9	Sun's gravity	7
													B	12	B	16	Universe's energy	10
													C	18	C	22	Heat from the sun	9
													D	*24*	*D*	*31*	Radioactivity	10
																	Not a valid response but chosen by 1 student on post	1

DISCUSSION

The purpose of this study was to answer the following questions:

1. How many GCI v. 1.0 questions showed the potential for conceptual entrenchment by having little, no, or negative change despite instruction?
2. Did these questions group in any way?
3. Did students show no change as a result of instruction because they are holding firmly to a belief (entrenchment) or because they are switching between conceptions (mobility)?

We found that 22 of the 73 GCI v. 1.0 questions had gains of <0.03 and that nearly half were basic physics and chemistry questions. We also discovered that students were far less likely to change answers on basic physics questions than they were for the geosciences ones, with many of the low-gain geoscience questions showing switch rates that were similar to the rate expected for guessing. In other words, the geosciences questions showed high conceptual mobility, whereas the physics conceptions appear to be more entrenched.

Previous studies have shown that, for many courses, little significant learning occurs across the test population as measured by the GCI v. 1.0 pre- to posttesting (Libarkin and Anderson, 2005), and our study shows that most student ideas about Earth are highly mobile for the lowest-gain questions. Although our work identifies these trends, we cannot at this point explain their origins. Although some students are undoubtedly guessing, the overall distribution of chosen distractors cannot be explained by guessing alone. In particular, physics questions as a group show the least mobility. These results lead to a number of questions regarding learning in the geosciences that warrant additional research.

Nearly half of our low-gain questions deal with basic physics and chemistry, prompting us to ask: Are students having difficulty understanding topics in Earth Science because of shaky supporting science underpinnings upon which geosciences concepts are built? A similar prevalence of, for example, gravity misconceptions in students enrolled in geoscience courses has been documented (Asghar and Libarkin, 2010). The lower switching rates for the physics questions suggests less conceptual mobility than for the geosciences concepts and perhaps a higher level of entrenchment, potentially preventing students from understanding geosciences concepts that require a solid a physical science foundation because of the short period over which

TABLE III: continued.

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Mountain Morphology	Changed to on Post1
30	0.03	48	12	13	9	39	81.3	0.32	8	7	5	58.3	A	8	A	13	Old mountain are taller because they grow	10
													B	25	B	28	Old mountains have gentler slopes - erosion	10
													C	15	C	16	Old mountains have more vegetation	10
				Average change in Group			74.0		Average change in Group			47.3	D	7	D	15	Old mountains are rougher because they crack	12
													E	5	E	4	All mountains are roughly the same age	5

Time

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Time Lines with Life	Changed to on Post1
28	-0.03	193	75	72	92	101	52.3	0.38	30	32	43	42.7	A	14	A	17	Life appears at Earth origin - rest correct	13
													B	19	B	17	Man and dino overlap	11
													C	68	C	64	Dino appears halfway through Earth History	32
													D	75	D	72	Correct	30
													E	17	E	16	Man and dino form at same time, early in Earth History	9

Pre
	A	B	C	D	E	NR
Post A	4	1	6	6	0	0
B	1	6	3	4	3	0
C	5	4	32	20	3	0
D	3	7	17	43	3	0
E	1	0	6	1	7	1
NR	0	1	4	1	4	1

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Techniques for Calculating the Age of the Earth	Changed to on Post1
17	0.03	392	17	27	50	342	87.2	0.96	20	10	7	58.8	A	250	A	235	Comparison of fossils found in rocks	57
													B	278	B	271	Comparison of different layers of rock	57
													C	128	C	185	Analysis of uranium and lead in rock	102
				Average change in Group			69.8		Average change in Group			50.7	D	231	D	221	Analysis of carbon in rock	50
													E	139	E	161	Measurement of erosion rates	75
													F	43	F	73	Measurement of the strength of the Earth's magnetic field	55
													G	96	G	99	Measurement of the height of mountains	64
													H	14	H	20	Scientists cannot calculate the age of the Earth	12

Atmospheric

GCI Question	Gain	Matching Exams	Pre-test Correct	Post-test Correct	No Pre to Post Change	Changed Pre to Post	% change	Rasch	To Correct	From Correct	Kept Correct	% change of pretest correct	Pre test		Post test		Where do most clouds form?	Changed to on Post1
49	-0.09	54	20	17	21	33	61.1	-0.56	9	12	8	60	A	3	A	5	1m2 land	4
													B	20	B	17	1m2 ocean	9
													C	13	C	8	1m2 plant covered	4
													D	18	D	24	1m2 equator	15

Pre
	A	B	C	D	NR
Post A	0	1	2	2	0
B	2	8	3	4	0
C	0	1	4	3	0
D	1	10	4	9	0
NR	0	0	0	0	0

Footnotes	
1	For Questions that allowed students to choose more than one answer, this number indicates answers chosen on post-test that were not chosen on pretest
2	Colors noted below can be viewed in the online article
	Blue represents number of students that did not change response pre to post
	Pink represents number of students that did not change the correct answer pre to post
	Green field represents correct answer
	Red represents a switch from one answer to another at a rate that was at least twice as high than for other switching possibilities.
	Gray represents a distractor that was preferentially chosen at a rate that was at least twice as high than for the other distractors
3	For questions allowing multiple correct answers, it is not possible to determine which answers students were switching from

entry-level courses are typically taught. Prior studies on entrenchment suggest that conceptual change requires periods longer than typical instruction for learning to occur (Vosniadou and Brewer, 1992). Little is known about the time needed for learning concepts built on ideas for which students do not already have a firm grasp, leading to the question: Is a semester enough time for students to develop a more accurate supporting science foundation and use this foundation to build accurate models of Earth phenomena? Do we need to pay more attention to basic physical science concepts in our introductory Earth Science courses, or require prerequisites, to provide students with a base upon which to build a solid understanding of Earth Science?

Even students who had highest pretest scores on the GCI struggled with some of these low-gain questions, similar to the finding by Libarkin and Anderson (2005) of insignificant overall gain on the GCI v. 1.0 for the highest pretesters. This suggests that over the course of a semester, even the best students have not showed significant learning as measured by the GCI. Yet, when the GCI is administered to advanced learners (graduate students and faculty), scores are high (Libarkin and Anderson, 2005). Clearly, a significant amount of learning as measured by the GCI is occurring sometime between the culmination of the introductory course and the upper- to graduate-level courses. When do the correct geosciences conceptions take root, what is the role of the introductory course in this later conceptual development, and what strategies can be employed in introductory courses to enhance learning for those students who will only take one college-level geosciences course? Can the completion of physics courses by geoscience majors, which generally occurs after introductory geoscience courses, help explain the gains in conceptual understanding that advanced students show? Are introductory geoscience courses necessary for laying a foundation upon which later learning can take place, or would a student who skipped the introductory course and entered the curricula at a more advanced stage learn equally well? Longitudinal studies of learning as measured by the GCI are critical in establishing a timeline upon which conceptual change occurs and may shed light on the role of introductory geoscience courses in learning and when our advance learners (majors) become proficient in their content knowledge. These studies should also inform us as to whether introductory geology courses are best viewed as a critical component for later advance-level learning or whether some latitude may be taken in the topics covered to better serve a general-education population without hurting the later development of our potential majors.

The geoscience question that students were most likely to change from a correct answer on the pretest to an incorrect answer on the posttest focused on the definition of a tectonic plate (87.5% of correct pretesters changed to wrong answers on the posttest, more than expected from guessing alone), and they were least likely to change their correct pretest answers on the posttest for three questions that dealt with the size and shape of Earth. We do not have interview data that sheds light on either of these observa-

tions, but the plate tectonic question is one of the more difficult questions on the GCI (Rasch score of 1.83, with only 8.8% of the pretest population answering correctly; Libarkin and Anderson, 2007b) and the size and shape of Earth are among the easiest (Rasch scores of −0.42, −0.67, and −1.16, with more than 40% of pretesters answering correctly). We speculate that students' lacking confidence in their understanding of the concepts at the root of difficult questions may compel them to switch from a correct answer.

Given the mobility that students show with respect to many Earth Science concepts, determination of the role and effect that K–12 Earth Science curricula have on students who eventually end up in our introductory college courses is warranted. Is it reasonable to send K–12 teachers into the classroom armed with a single-semester introductory Earth Science course if little learning occurred over this period? Should K–12 districts that presently teach Earth Science in 8th or 9th grade, and significantly before physics and chemistry, rethink the ordering of the various science courses in their middle school and high school curricula? Dahl et al. (2005) published data on K–12 teacher interest that ranked plate tectonics last in a list of Earth Science concepts, clearly different from how many geologists view what is perhaps the major unifying concept in our field (e.g., Earth Science Literacy Initiative, http://www.earthscienceliteracy.org; Dahl et al., 2005). This illustrates a major disconnect between geology professionals and those responsible for laying a conceptual Earth Science foundation for eventual college students, and it demonstrates the challenge geosciences educators face in conducting research on conceptual understanding that will inform strategies for bridging this gap between K–12 geoscience preparation and college-level expectations of learning.

Our study is enhanced by the large number of students who completed the pre- and posttests. However, a large sample size also leads to limitations in terms of understanding some of the reasons for the trends that we find in the data. One of the biggest drawbacks of our study is that we lack in-depth data on participants' backgrounds and demographics because of time constraints during test administration. We specifically desire additional information on students' science preparation in high school, their knowledge of basic chemistry and physics principles, their attitudes toward science and scientists that may affect their ability to learn the material, and their motivation for enrolling in an introductory Earth Science course. Interviews of representative students that focus on their pre-college science background and experiences may be needed to provide a more complete assessment of the trends outlined in this study.

Many of the questions and preliminary conclusions presented here require additional study. The observation that students with the highest pretest scores show no significant improvement on the GCI, yet graduate students and professors have nearly perfect GCI scores, suggests that learning as measured by the GCI occurs sometime after completing the introductory course and before graduation with a geoscience degree. Longitudinal studies that use the GCI to follow individual students through each stage or semester of their undergraduate geoscience training should pinpoint where learning gains occur. We can then assess whether these gains correspond to the completion of any particular geoscience or supporting science courses or whether there is a slow and steady improvement of scores as students immerse themselves in the field of study and have time to blend together their geologic and supporting science information to form the more complex conceptual understanding needed to improve on the GCI.

In addition, it may be necessary to incorporate surveys that assess the affective domain to better understand the role of motivations and attitudes and how they link to changes in learning as measured by the GCI. Also, we do not understand the variation in pretest GCI scores and why some introductory students have low geoscience understanding, whereas others exhibit a much higher level of understanding. Did the high pretesters complete specific geoscience courses in high school, or did they simply have a better supporting science background before entering college? Do they have better quantitative backgrounds? Do they exhibit different levels of motivation and attitudes toward science? A more in-depth survey of student pretester backgrounds could shed light on what it takes to properly prepare students for our college geoscience courses.

REFERENCES

Adams, R.J., and Khoo, S. 1996. *Quest: The Interactive Test Analysis System, Version 2.1*. Camberwell, Australia: Australian Council for Educational Research.

Allen, K. 2006. The statistics concept inventory: Development and analysis of a cognitive assessment instrument in statistics [unpublished dissertation]. Norman: The University of Oklahoma.

Anderson, D.L., Fisher, K.M., and Norman, G.L. 2002. Development and validation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching,* 39:952–978.

Asghar, A.A., and Libarkin, J.C. 2010. Gravity, magnetism, and "down": Non-physics college students' conceptions of gravity. *Science Educator,* 19(1):42–55.

Bruer, J.T. 1993. The mind's journey from novice to expert. *American Educator,* 17:6–46.

Chi, M.T.H., Sotta, J.D., and de Leeuw, N. 1994. From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction,* 4:27–43.

Dahl, J., Anderson, S.W., and Libarkin, J.C. 2005. Digging into Earth Science: Alternative conceptions held by K–12 teachers. *Journal of Science Education,* 12:65–68.

Gardner, H. 1991. The unschooled mind. New York: Basic.

Gery, F.W. 1972. Does mathematics matter? *In* Welch, A., ed., Research papers in economic education. New York, NY: Joint Council on Economic Education, p. 142–157.

Gordon, C. 1992a. A case study of conceptual change. *In* Kinzer, C.K., and Leu, D.J., eds., Literacy research, theory and practice: Views from many perspectives. Presented at the Forty-first yearbook of the National Reading Conference, Chicago, IL, p. 161–168.

Gordon, C. 1992b. A case study of conceptual change: a follow-up. Presented at the National Reading Conference, San Antonio, Texas.

Hake, R.R. 1997. Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics,* 66:64–74.

Halloun, I.A., and Hestenes, D. 1985a. Common sense concepts about motion. *American Journal of Physics,* 53(11):1056–1065.

Halloun, I.A., and Hestenes, D. 1985b. The initial knowledge state of college physics students. *American Journal of Physics,* 53(11):1043–1055.

Harrison, A.G., Grayson, D.J., and Treagust, D.F. 1999. Investiga-

tion of a grade 11 student's evolving conceptions of heat and temperature. *Journal of Research in Science Teaching,* 36:55–87.

Hestenes, D., Wells, M., and Swackhamer, G. 1992. Force Concept Inventory. *Physics Teacher,* 30:141–158.

Hovland, C.I., Lumsdaine, A.A., and Sheffield, F.D. 1949. A baseline for measurement of percentage change. *In* Hovland, C.I., Lumsdaine, A.A., and Sheffield, F.D., eds., Experiments on mass communication. Princeton, NJ: Princeton Universtiy Press, p. 284–289.

Johnson, S.D., and Roellke, C.F. 1999. Secondary teachers' and undergraduate education faculty members' perceptions of teaching effectiveness criteria: A national survey. *Communication Education,* 48:127–138.

Krause, S., Birk, J.P., Bauer, R.C., Jenkins, B., and Pavelich, M.J. 2004. Development, testing, and application of a chemistry concept inventory. Proceedings of the Frontiers in Education Conference, FIE 2004. 34th Annual. IEEE: Piscataway, NJ, p. T1G–1.

Libarkin, J.C. 2006. College student conceptions of geological phenomena and their importance in classroom instruction. *Planet,* 17:6–9.

Libarkin, J.C. 2008. Concept inventories in higher education science. *In* National Research Council Promising Practices in Undergraduate STEM Education Workshop 2. Washington, DC: National Academies Press.

Libarkin, J.C., and Anderson, S.W. 2005. Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education,* 53:394–401.

Libarkin, J.C., and Anderson, S.W. 2007a. Development of the Geoscience Concept Inventory. *In* Proceedings of the National STEM Assessment Conference. Washington DC: National Science Foundation, October 19–21, 2006, p. 148–158.

Libarkin, J.C., and Anderson, S.W. 2007b. The Geoscience Concept Inventory: Application of Rasch analysis to concept inventory development in higher education. *In* Liu, X., and Boone, W., eds., Applications of Rasch Measurement in Science Education. Maple Grove, MN: JAM Publishers, p. 45–73.

Libarkin, J.C., Anderson, S.W., Dahl, J., Beilfuss, M., Boone, W., and Kurdziel, J.P. 2005. Qualitative analysis of college students' ideas about the Earth: Interviews and open-ended questionnaires. *Journal of Geoscience Education,* 53:17–26.

Libarkin, J.C., Ward, E.M.G., Anderson, S.W., Kortemeyer, G., and Raeburn, S. 2011. Revisiting the Geoscience Concept Inventory: A call to the community. *GSA Today,* 21:26–28.

Lindell, R.S., and Olsen, J.P. 2002. Developing the lunar phases concept inventory. Proceedings of the 2002 Physics Education Research Conference. New York: Perc Publishing.

Maria, K. 1997. A case study of conceptual change in a young child. *Elementary School Journal,* 22:1–35.

Reed-Rhoads, T. 2008. Concept inventories in engineering education. Commissioned paper for workshop on Linking Evidence and Promising Practices in Science, Technology, Engineering, and Mathematics (STEM) Undergraduate Education. Washington, DC: National Science Foundation.

Thornton, R.K., and Sokoloff, D.R. 1998. Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics,* 66:338–352.

Vosniadou, S., and Brewer, W.F. 1992. Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology,* 24:535–585.

Yeo, S., and Zadnik, M. 2001. Introductory thermal concept evaluation. *Physics Teacher* 39:496–503.

Zeilik, M., Schau, C., and Mattern, N. 1999. Conceptual astronomy. II. Replicating conceptual gains, probing attitude changes across three semesters. *American Journal of Physics,* 67(10):923–972.